CLASSIFICATION INTO TWO MULTIVARIATE NORMAL POPULATIONS WITH KNOWN MEANS

by

Somesh Das Gupta
University of Minnesota

Technical Report No. #368
February, 1980

1. Introduction

We consider the problem of classifying an (randomly observed) experimental unit ω into one of two populations Π_1 and Π_2. It is assumed that a random vector $X(\omega) : p \times 1$ is distributed as $N_p(\mu_i, \Sigma)$ in Π_i ; furthermore, μ_1 and μ_2 are assumed to be known, but Σ is an unknown postive-definite matrix. Based on random training samples from Π_1 and Π_2 we may get a statistic $S : p \times p$, where nS is distributed as the Wishart distribution $\mathcal{W}_p(n;\Sigma)$, $n \geq p$.

If Σ were known, the admissible Bayes and minimax rule δ classifies ω into Π_1 iff

$$(1) \qquad (x - \mu_1)'\Sigma^{-1}(x - \mu_1) < (x - \mu_2)'\Sigma^{-1}(x - \mu_2) .$$

The plug-in rule $\hat{\delta}$ (obtained by replacing Σ by S) classifies ω into Π_1 iff

$$(2) \qquad T \equiv (\mu_2 - \mu_1)'S^{-1}(x - [\mu_1 + \mu_2]/2) < 0 .$$

The likelihood-ratio rule suggested by Anderson [1] also turns out to be the same as $\hat{\delta}$. In this note we shall study the rule $\hat{\delta}$.

2. Probabilities of Correct Classification

Let

$$(3) \qquad \Delta = [(\mu_1 - \mu_2)'\Sigma^{-1}(\mu_1 - \mu_2)]^{\frac{1}{2}}$$

Without lost of generality, we may assume that

$$(4) \qquad \mu_1 = 0 \ , \ \mu_2 = (\Delta, 0, \ldots, 0)' \ , \ \Sigma = I_p \ .$$

Let $A = nS = [a_{ij}]$, $A^{-1} = [a^{ij}]$. Then it can be seen that

(5) $\quad T/n = \Delta[x_{1.(2)} - \Delta/2]/A_{11.(2)}$;

where

(6) $$A = \begin{array}{cc} \left[\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{12} \end{array}\right] & \begin{array}{c} 1 \\ p-1 \end{array} \\ \begin{array}{cc} 1 & p-1 \end{array} & \end{array} , \quad A_{11.2} = A_{11} - A_{12}A_{22}^{-1}A_{21} ,$$

(7) $$X_{1\ (2)} = X_1 - A_{12}A_{22}^{-1}X_{(2)} , \quad X = (X_1, X_2, \ldots, X_p)' = (X_1, X'_{(2)})' .$$

The conditional distribution of $[X_{1.(2)} - \Delta/2]$, given $X'_{(2)}A_{22}^{-1}X_{(2)} = v$, is given by

(8) $$N((-1)^i\Delta/2 \ , \ 1 + v)$$

independent of $A_{11.(2)}$, if ω comes from Π_i . Note that $X'_{(2)}A_{22}^{-1}X_{(2)}$ and $A_{11.(2)}$ are distributed independently as $f_{p-1,\ n-p+2}$ and χ^2_{n-p+1} , respectively, where $f_{a,b}$ is the distribution of the ratio of independent χ^2_a and χ^2_b . Thus the probabilities of correct classification are given by

$$P(T \leq 0|\Pi_1) = P(T \geq 0|\Pi_2) \ .$$

(9) $$= E\ \Phi\ (\Delta\ \beta^{\frac{1}{2}}/2) \ ,$$

where Φ is the c.d.f. of $N(0, 1)$ and $\beta^{\frac{1}{2}}$ is distributed as the square root of a $\beta(\frac{n-p+2}{2}, \frac{p-1}{2})$ variate.

3. Optimum properties of $\hat{\delta}$

Consider an orthogonal matrix $L : pxp$ such that the first row is proportional to $(\mu_2 - \mu_1)'$. Define

(10) $$Y = L(x - \mu_1) \ , \ B = LAL' = [b_{ij}] \ , \ \Gamma = L\Sigma L'$$

Then $Y \sim N_p(0,\Gamma)$ in Π_1 and $Y \sim N_p(\nu, \Gamma)$ in Π_2, where

(11) $\nu = (c, 0, \ldots, 0)'$, $c = [(\mu_1 - \mu_2)'(\mu_1 - \mu_2)]^{\frac{1}{2}}$

Consider now the classification problem in terms of Y and B.

The problem is invariant under transformations:

(12) $Y \to TY$, $B \to TBT'$,

where

$$(13) \qquad T = \begin{array}{cc} \left[\begin{array}{c|c} 1 & t'_{(1)} \\ \hline 0 & T_{22} \end{array}\right] & \begin{array}{c} 1 \\ p-1 \end{array} \\ \begin{array}{cc} 1 & \quad p-1 \end{array} & \end{array}$$

is a nonsingular matrix. A set of maximal invariant statistics is given by

(14) $U = Y_{1.(2)}$, $V = Y'_{(2)} B_{22}^{-1} Y_{(2)}$, $b_{11.(2)}$,

where

$$(15) \qquad B = \begin{array}{cc} \left[\begin{array}{c|c} b_{11} & B_{12} \\ \hline B_{21} & B_{22} \end{array}\right] & \begin{array}{c} 1 \\ p-1 \end{array} \\ \begin{array}{cc} 1 & \quad p-1 \end{array} & \end{array} , \quad b_{11.(2)} = b_{11} - B_{12} B_{22}^{-1} B_{21} ,$$

(16) $Y_{1.(2)} = Y_1 - B_{12} B_{22}^{-1} Y_{(2)}$, $Y = (Y_1, Y_2, \ldots, Y_p)' = (Y_1\ Y'_{(2)})'$

Note that $b_{11.(2)}$ is independent of U and V, and

$b_{11.(2)} \sim \gamma_{11.(2)} \chi^2_{n-p+1}$, where

$$(17) \qquad \gamma_{11.(2)} = \Gamma_{11} - \Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21} , \qquad \Gamma = \begin{array}{cc} \left[\begin{array}{c|c} \gamma_{11} & \Gamma_{12} \\ \hline \Gamma_{21} & \Gamma_{22} \end{array}\right] & \begin{array}{c} 1 \\ p-1 \end{array} \\ \begin{array}{cc} 1 & \quad p-1 \end{array} & \end{array}$$

Note that

(18) $$\Delta^2 = (\mu_1 - \mu_2)'\Sigma^{-1}(\mu_1 - \mu_2) = \nu' \Gamma^{-1} \nu = c^2/\gamma_{11.(2)}$$

Hence

(19) $$\gamma_{11.(2)} = c^2/\Delta^2 .$$

Also

$$(\mu_2 - \mu_1)\ A^{-1}(X - [\mu_1 + \mu_2]/2)$$

(20) $$= \nu' B^{-1}[Y - \nu/2] = c(Y_{1.(2)} - c/2)/b_{11.(2)} .$$

The distribution of U, given $V = v$, is given by

(21) $$N(\theta, \gamma_{11.(2)}(1 + v)) ,$$

where $\theta = 0$ in Π_1 and $\theta = c$ in Π_2 .

Consider a prior distribution which sets $\gamma_{11.(2)} = \gamma_0$ and attaches equal probabilities to Π_1 and Π_2. Then the unique (a.e.) Bayes rule (with zero-one loss function) in the class of all invariant rules classifies ω into Π_1 iff

(22) $$U < c/2 .$$

Thus this invariant admissible Bayes rule coincides with $\hat{\delta}$. Since for a fixed Δ the probabilities of correct classification for $\hat{\delta}$ are equal, it can be easily seen that $\hat{\delta}$ is invariant minimax.

Kiefer and Schwarz [3] derived a class of Bayes rules (not invariant) for this problem. Such a Bayes rule decides ω in Π_1 iff

$$\mathrm{etr}[(\mu_2 - \mu_1)'X] \quad [1 + (X - \mu_1)'A^{-1}(x - \mu_1)]^{-r/2} \quad [1 + (X - \mu_2)'A^{-1}(X - \mu_2)]^{r/2}$$

(23) $$> k .$$

The rule $\hat{\delta}$ does not provide a similar test if this decision problem is viewed as a hypothesis testing problem. Such a similar region (for ω in Π_1) is given by

(24) $$\sqrt{n-p+1}\ Y_{1.(2)}/[(1 + Y'_{(2)}\ B_{22}^{-1}\ Y_{(2)})\ b_{11.(2)}]^{\frac{1}{2}} < k ,$$

since the statistics in the left-hand side of (24) is distributed as Student's $t_{n-p} + 1$.

Note: The above problem was also considered by S. Geisser [2] from the viewpoint of predictive discrimination; he used an improper prior distribution for Σ.

References

1. Anderson T. W. (1958). An Introduction to Multivariate Statistical Analysis. Wiley, New York.

2. Geisser, S. (1966). Predictive Discrimination. Multivariate Analysis, I; ed. P. R. Krishnaiah. Academic Press, N.Y.

3. Kiefer, J. and Schwartz, R. (1965). Admissible Bayes character of T^2 — , R^2 — , and Other Fully Invariant Tests for Classical Multivariate Normal Problems. Ann. Math. Statist., 36, 747-770.